



Research article

Image classification of Chinese medicinal flowers based on convolutional neural network

Meiling Huang* and Yixuan Xu

Department of Industrial Engineering & Management, National Chin-Yi University of Technology, 57, Sec. 2, Chung-Shan Rd., Taiping, Taichung, Taiwan

* **Correspondence:** Email: huangml@ncut.edu.tw; Tel: +886423924505; Fax: +886423934620.

Abstract: *Background and objective:* Traditional Chinese medicine has used many herbs on the prevention and treatment of diseases for thousands of years. However, many flowers are poisonous and only few herbs have medicinal properties. Relying on experts for herbs identification is time consuming. An efficient and fast identification method is proposed in this study. *Methods:* This study proposes ResNet101 models by combining SENet and ResNet101, adding convolutional block attention module or using Bayesian optimization on Chinese medicinal flower classification. The performances of the proposed ResNet101 models were compared. *Results:* The best performance for accuracy, precision, recall, F1-score and PR-AUC are coming from ResNet101 model with Bayesian optimization which are 97.64%, 97.99%, 97.86%, 97.82% and 99.72%, respectively. *Conclusions:* The proposed ResNet101 model provides a better solution on the image classification of Chinese medical flowers with favourable accuracy.

Keywords: Chinese medicinal flower; Squeeze-and-Excitation; classification; deep learning

1. Introduction

Chinese medicinal flowers are very important in traditional Chinese medicine. The pharmacology of Chinese medicinal flowers is widely used in the prevention and treatment of various diseases. It has been an indispensable medicinal tool for thousands of years [1,2]. Tree flowers are one of the important categories of traditional Chinese medicine and serve not only as food but also as herbal medicine to treat diseases. The variety and complexity of traditional Chinese medicine flowers are many and many

medicinal plants look alike and confusing. Only experienced experts can identify and classify the tree flowers of Chinese medicine accurately. However, relying on visual assessment from experts for herb identification is time consuming and subjective [3]. It is suggested to develop an efficient and fast identification model to identify flowers correctly and prevent herb poisonings [4].

Common identification methods on traditional Chinese medicinal materials use DNA barcodes, FT-IR, SD-IR and 2D-IR biochemical identification [5,6]. Nowadays, deep learning recognizes images automatically and effectively and convolutional neural networks have achieved favorable results in image classification. Jahanbakhshi et al. used Multilayer perceptron (MLP), Fuzzy, k-nearest neighbor (KNN), support vector machine (SVM), gradient boosting tree (GBT) and evolution decision tree (EDT) to classify the extracted features [7]. The feature extraction of traditional machine learning mainly relies on manual labor while deep learning automatically extracts features from images. Convolutional neural networks have achieved good results in the classification research of flower recognition [8] and traditional Chinese medicine recognition [2]. The related literatures using convolutional neural networks on the classification of multiple traditional Chinese medicines are few. Chai et al. [9] and Jahanbakhshi et al. [7] classify a single type and its counterfeits. There are also some literatures that classify multiple types of Chinese medicines. For example, Xu et al. [10] created a NH-98 Chinese medicine dataset by themselves and used multiple Chinese medicines for classification prediction. However, an unbalanced dataset may affect the research results.

In addition, all of the related Chinese medicine datasets are self-photographed or collected datasets from the internet. Currently, there is no open public dataset of Chinese medicine flowers available. We selected twelve most commonly used and economically valuable Chinese medical flowers and created a dataset of Chinese medicinal flowers in previous study [11]. The dataset provides a collection of blossom images both close-up photos and remote photos.

ResNet-101 is a deep residual neural network [12] developed by researchers at Microsoft. It is a residual network. The architecture of ResNet-101 is a variant of the ResNet series of models. Compared to the shallower architecture (for example: ResNet-50), it has more layers. The residual learning proposed by ResNet makes the network deeper. It is easier to train and can solve the problem of gradient disappearance (Vanishing Gradient). Since ResNet-101 has deeper layers, it has stronger modeling ability and can capture more complex features and patterns. This makes it perform better in many computer vision tasks such as image classification, object detection [13] and semantic segmentation [13]. Due to better initialization and better gradient flow, ResNet-101 can often converge faster relative to shallower models. This enables the model to achieve better performance in less time.

SENet is an attention mechanism module to improve network performance. Many literatures add attention modules to the convolutional neural network model [14,15]. Neural architecture search (NAS) designs the architectures automatically and has been applied in related works [16–19]. A comprehensive review of the NAS algorithm was done to summarize the design principles and enlighten the future direction [20]. At present, SENet has been flexibly applied to existing network architectures such as MobileNet and ResNet50 [21]. At the same time, SENet is also applicable to other networks. He et al. [22] proposed a YOLOv4-based model Mina-Net in 2022 to detect insulator explosion images by using channel attention mechanism to improve the recognition accuracy of the network. The results show that the accuracy of Mina-Net is 88.07% which can effectively improve the detection accuracy of self-explosion of different sizes which is 4.78% higher than that of YOLOv4. Yang et al. [23] proposed a DAN-EffcinetNet-B2 model for fish feeding behavior recognition in 2021 with a dual attention network of EffcinetNet-B2. The output feature vectors of the last layer are sent

to channel attention and spatial attention respectively. Finally, the outputs of the two attention modules are fused which can focus on the interdependence of space and channels and solve the feature extraction problem of fish gathering areas during feeding. The experimental results show that the test accuracy is 89.56% which is higher than results from AlexNet, VGG, InceptionV3 and ResNet. Zhang et al. [24] proposed an aggressive posterior retinopathy of prematurity (AP-ROP) in 2022, adding a channel attention module and a bilinear pooling module to obtain complementary information between layers. The results show that the accuracy of the combined network is 95.81% which is better than ResNet50 and ResNet101.

Zhou et al. [14] proposed a model for automatic ore classification in 2022. After data augmentation and transfer learning, MobileNet model has an accuracy of 94% while the accuracy reached 96.89% when adding SENet to MobileNet. Zhao et al. [15] developed a vegetable disease recognition model DTL-SE-ResNet50 in 2022, the results show that the recognition accuracy is 97.24% which is better than models such as EfficientNet, AlexNet, VGG19 and InceptionV3.

This study proposes SE-ResNet101 model. To our understanding, this is the first study combining SENet and ResNet101 on Chinese medicinal flower classification. We compare the performances of the proposed SE-ResNet101 models with four models including AlexNet, InceptionV3, ResNet50 and ResNet101. In addition, the CBAM-ResNet101 and SE-CBAM-ResNet101 models are conducted in this study to observe if there is synergy between the different changes. The main contributions of the study are as follows:

- 1) Develop classification models for the 12 most commonly used Chinese medicinal flowers.
- 2) The first study combines SENet with ResNet101 models on classification of Chinese medicinal flowers.
- 3) Compare the performances on connection of attention mechanism and ResNet101 models on classification of Chinese medicinal flowers.

The remainder of this paper is structured as follows. Section 2 introduces the dataset used in this research and the proposed models in this study. Section 3 presents and compares the results of the proposed models with other models. Section 4 presents the ablation experiments. The conclusion of this study is made in Section 5.

2. Materials and methods

2.1. Chinese medicinal blossom-dataset

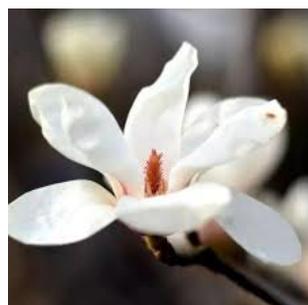
The dataset of blossom images for traditional Chinese medicinal flowers were searched through the web. There are twelve categories: 1) *Syringa*, 2) *Bombax malabarica*, 3) *Michelia alba*, 4) *Armeniaca mume*, 5) *Albizia julibrissin*, 6) *Pinus massoniana*, 7) *Eriobotrya japonica*, 8) *Styphnolobium japonicum*, 9) *Prunus persica*, 10) *Firmiana simplex*, 11) *Ficus religiosa* and 12) *Areca catechu*. The total number of searched images is 1716. The sub-folders are named by blossom categories in Mendeley. Figures 1 and 2 display examples of the close-up photos and remote photos for twelve categories.



1) Syringa



2) Bombax malabarica



3) Michelia alb



4) Armeniaca mume



5) Albizia julibrissin



6) Pinus massoniana



7) Eriobotrya japonica



8) Styphnolobium japonicum



9) Prunus persica



10) Firmiana simplex



11) Ficus religiosa

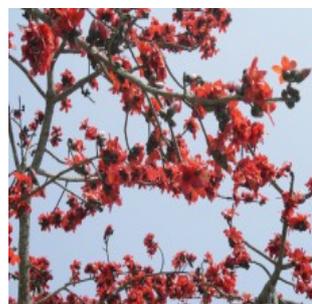


12) Areca catechu

Figure 1. Examples of the close-up photos for twelve categories.



1) Syringa



2) Bombax malabarica



3) Michelia alba



4) Armeniaca mume

Continued on next page

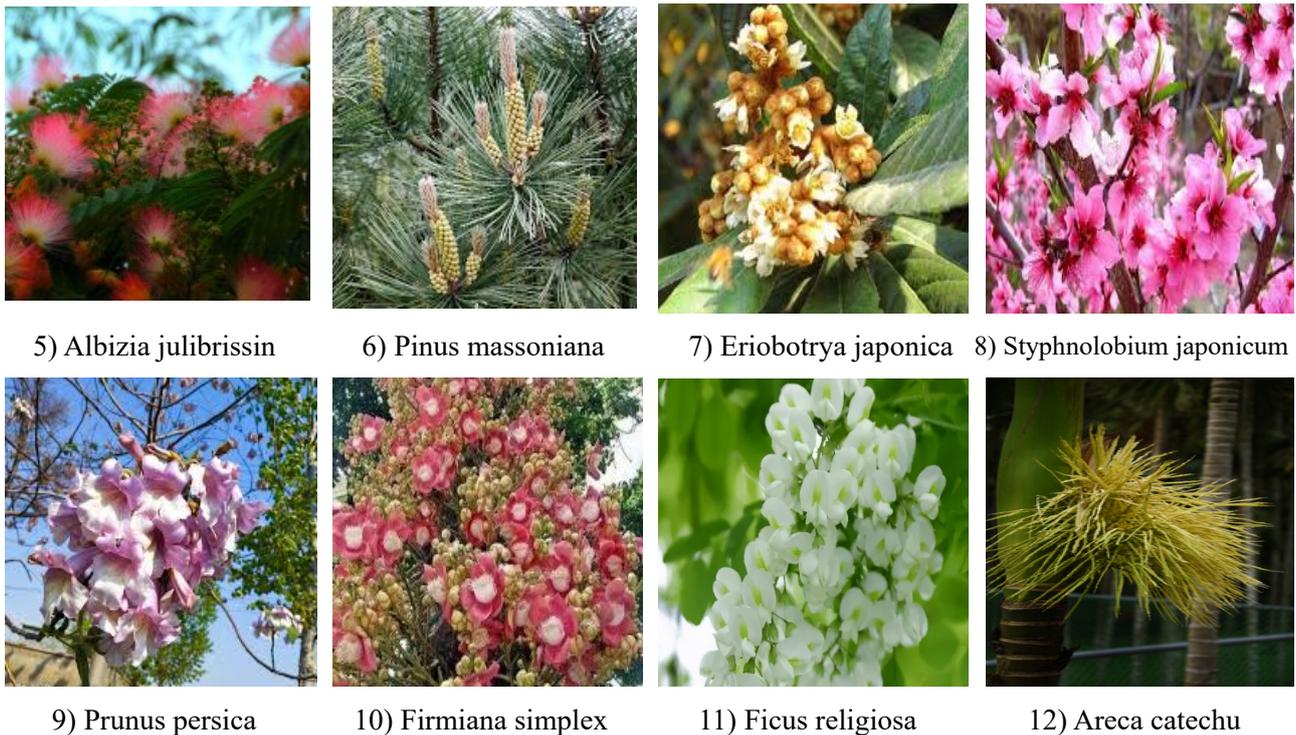


Figure 2. Examples of the remote photos for twelve categories.

2.1.1. Image preprocessing

Some original images quality are poor. We crop letters and frames, remove handwriting and blurry parts, center flowers, and adjust length and width, etc. Before data augmentation, the numbers of images for each category are 1) *Syringa*, 191; 2) *Bombax malabarica*, 172; 3) *Michelia alba*, 122; 4) *Armeniaca mume*, 236; 5) *Albizia julibrissin*, 222; 6) *Pinus massoniana*, 87; 7) *Eriobotrya japonica*, 115; 8) *Styphnolobium japonicum*, 213; 9) *Prunus persica*, 89; 10) *Firmiana simplex*, 75; 11) *Ficus religiosa*, 126; 12) *Areca catechu*, 68.

2.1.2. Image augmentation

Data augmentation enhances classification performance and stability of models through creating image diversity. Data augmentation methods are many and we selected Gaussian filtering, image brightness augmentation, image brightness reduction, mirror rotation, noise increase, 90° and 180° rotation methods, seven methods in total in the training and validation datasets. After that, all the images are augmented through mirror flip. Finally, dataset was increased to eight times. Figure 3 shows an example of the original image and the images after data augmentation. Table 1 presents the number of training, validation and test images before and after data augmentation.

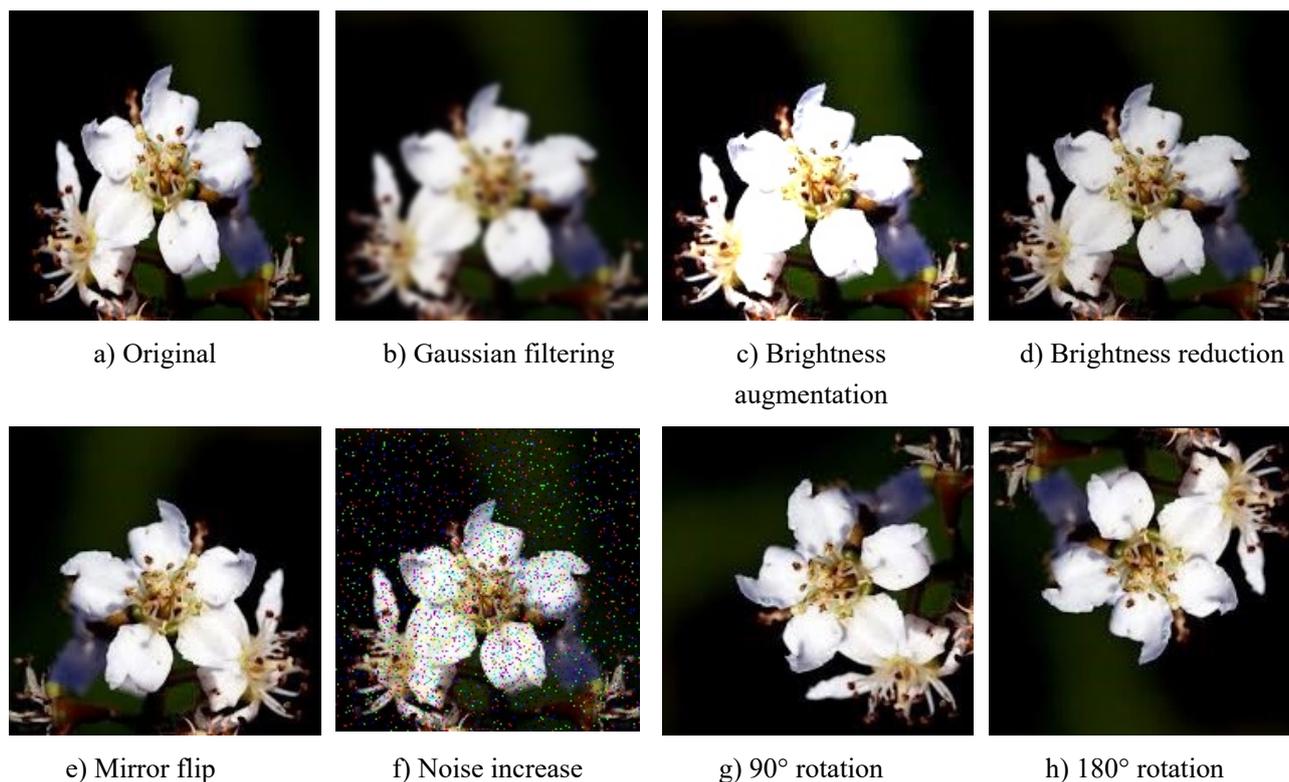


Figure 3. Example of data augmentation.

Table 1. Number of images before and after data augmentation.

ID	Name	Original				After data augmentation			
		Train	Val	Test	Total	Train	Val	Test	Total
1	Syringa	153	19	19	191	1224	152	19	1395
2	Bombax malabarica	138	17	17	172	1104	136	17	1257
3	Michelia alba	98	12	12	122	784	96	12	892
4	Armeniaca mume	188	24	24	236	1504	192	24	1720
5	Albizia julibrissin	178	22	22	222	1424	176	22	1622
6	Pinus massoniana	70	9	8	87	560	72	8	640
7	Eriobotrya japonica	92	11	12	115	736	88	12	836
8	Prunus persica	171	21	21	213	1368	168	21	1557
9	Firmiana simplex	72	9	8	89	576	72	8	656
10	Ficus religiosa	60	7	8	75	480	56	8	544
11	Styphnolobium japonicum	101	13	12	126	808	104	12	924
12	Areca catechu	55	6	7	68	440	48	7	495
Total		1376	170	170	1716	11,008	1360	170	12,538

2.2. Research methodologies

Figure 4 displays the flowchart of this study.

Step 1: The blossom images of traditional Chinese medicinal flowers were searched and created through the web. There are both close-up photos and remote photos for each category.

Step 2: Divide the original images into training set, validation set and test set at 80:10:10 ratio.

Step 3: Increase the number of images to eight times through data augmentation.

Step 4: Apply several classification models including AlexNet, InceptionV3, ResNet50 and ResNet101.

Step 5: Compare the performance of the above classification models.

Step 6: Evaluate the ablation experiments of the proposed ResNet101 models on Chinese medicinal flower dataset.

Step 7: Compare the performances of the proposed ResNet101 models with VGG and DenseNet models on Chinese medicinal flower dataset.

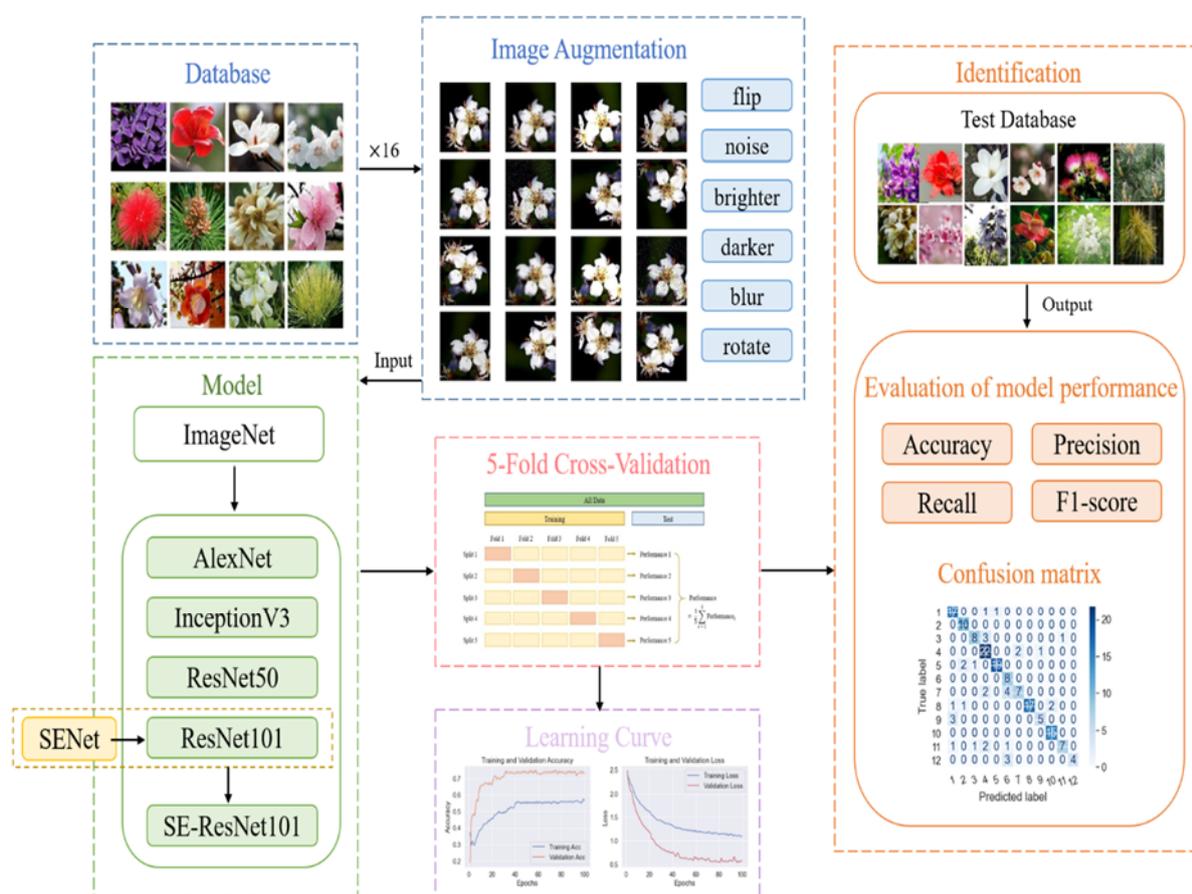


Figure 4. Flowchart of this study.

2.3. The proposed SE-ResNet101 model

The more layers of neural network, the more complex the architecture. residual network (ResNet) proposes a residual block with shortcut connections which solves the degradation problem of deep networks, accelerates network training and conquers the gradient disappearance and gradient explosion

as the number of network layers increases in the model. ResNet101 is a commonly used convolutional neural network. As the name implies, it is a deep convolutional neural network with 101 layers. The network can be divided into five parts: conv1, conv2, conv3, conv4, and conv5. Figure 5 details the architecture of ResNet101 model.

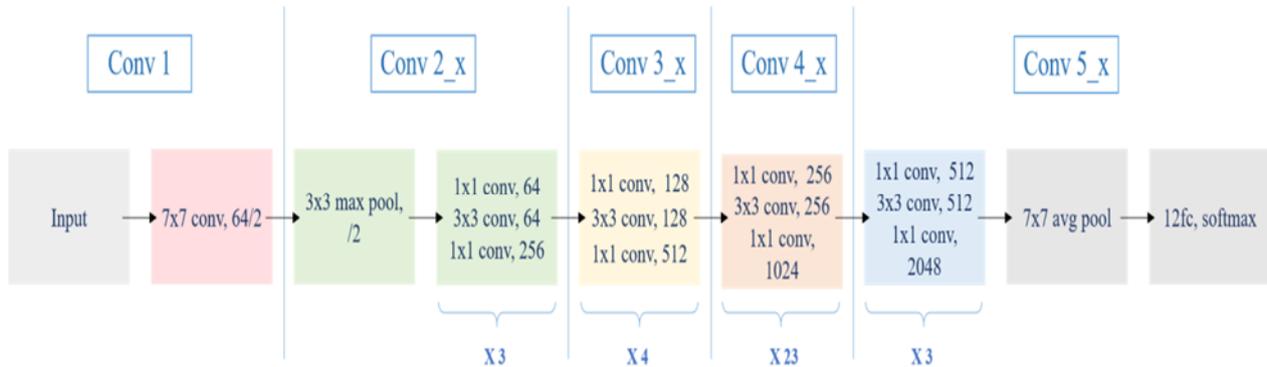


Figure 5. Architecture of ResNet101.

Squeeze-and-excitation network (SENet) is a representative channel attention module with simple structure, easy deployment, no new functions or layers [25], and easily combined with the existing network. SENet reduces the error rate of the model, and has low complexity, low calculation amount. Figure 6 presents the architecture of SENet where C , H and W represent the number of image channels, image height and image width, respectively.

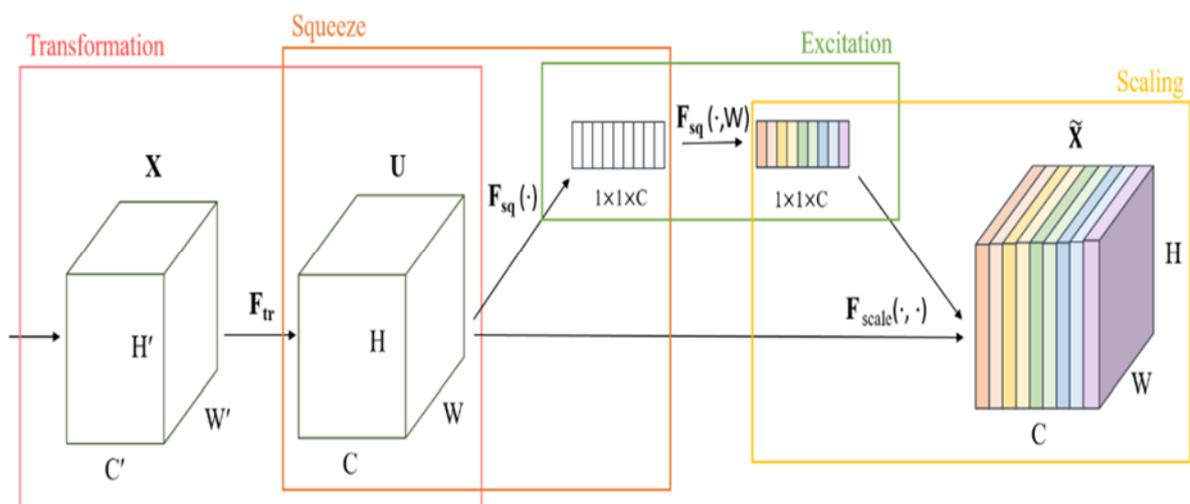


Figure 6. Architecture of SENet.

Transformation: F_{tr} represent the transformation relationship from X to U .

$$F_{tr}: X \rightarrow U, X \in \mathbb{R}^{H' \times W' \times C'}, U \in \mathbb{R}^{H \times W \times C} \quad (1)$$

Squeeze: squeeze H and W to one dimension and transform the input of $H \times W \times C$ to $1 \times 1 \times C$.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \text{ where } z \in \mathbb{R}^C \quad (2)$$

Excitation: The dimension of s is $1 \times 1 \times C$; C represents the number of channels.

$$s = F_{sq}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

Scaling: The multiplication between s and U to scale the weights. u_c is a two dimension matrix; s_c represents the weight.

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (4)$$

The proposed SE-ResNet101 model adds SENet to pre-trained ResNet101 on ImageNet after the last global average pooling layer (GlobalAveragePooling) to achieve better performance. Figure 7 presents the present the layer-by-layer architecture of the proposed SE-ResNet101 model used in this study and the proposed model is detailed as follows:

- 1) Adding SENet after the last global average pooling layer (GlobalAveragePooling)
- 2) Adding one more global average pooling layer and dense layer after SENet
- 3) Using softmax
- 4) Adding Dropout at 0.5.

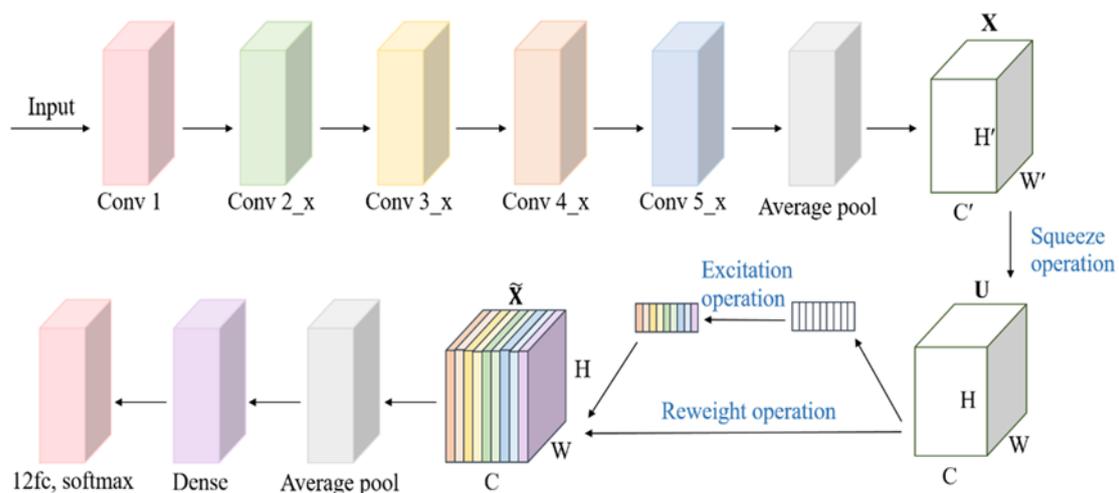


Figure 7. Architecture of the proposed SE- ResNet101.

2.4. The proposed CBAM-ResNet101 model

The convolutional block attention module (CBAM) combines channel and spatial attention mechanisms and has been applied to many common convolutional neural networks and it has been confirmed that it can effectively improve the performance of convolutional neural networks for image classification, target detection and other tasks. The CBAM module is mainly divided into a channel attention module and a spatial attention module. The details of the two modules will be explained below [25].

1) Channel attention module: first input the feature map to the global maximum pooling and global average pooling layers. Then, after processing by the shared Multilayer perceptron multi-layer perceptron add the two results and operate through the sigmoid activation function to generate each Channel weight values. Finally, multiply each channel weight value with the input feature map.

2) Spatial attention module: Take the maximum value and average value of each feature point of the feature map output by the channel attention mechanism to stack. Then, use the convolutional layer with a channel number of 1 to reduce the dimension and operate through the sigmoid activation function to generate Spatial attention feature map. Finally, multiply this feature map with the input feature map to get the final feature map.

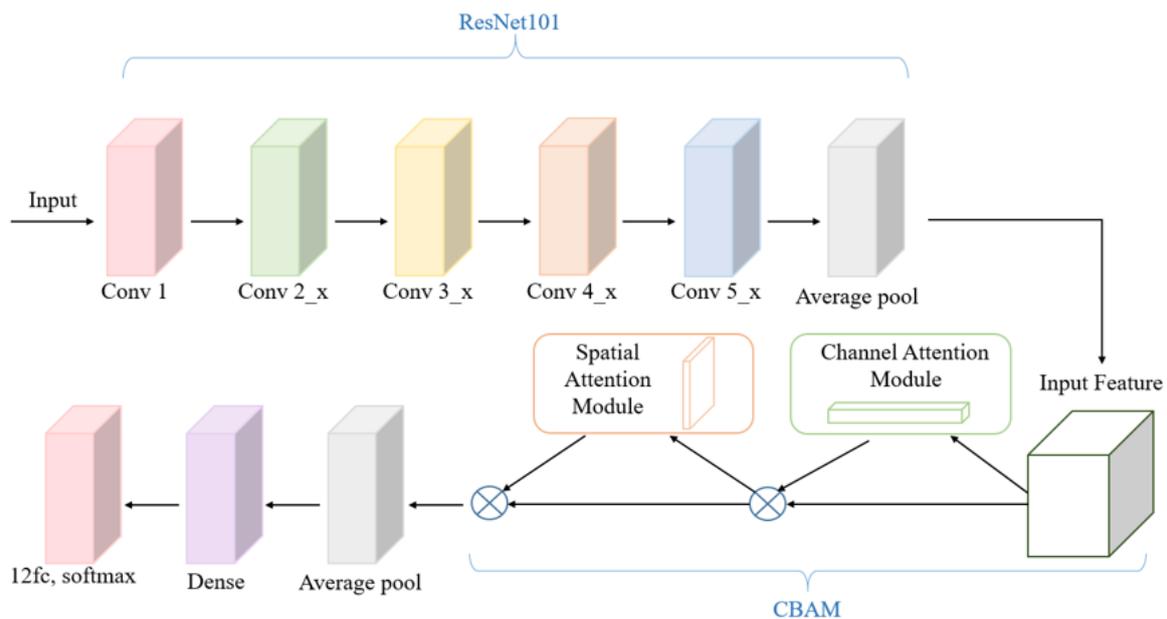


Figure 8. Architecture of the proposed CBAM-ResNet101.

Figure 8 presents the architecture of the proposed CBAM-ResNet101 model. Adding a CBAM module after the last average pooling layer of ResNet101 can retain more effective features than the original ResNet101. Compared with the SE module, CBAM contains channel attention mechanism and spatial attention mechanism. CBAM combines two attention mechanisms can capture the important features more effectively in the image. The channel attention module (CAM) adaptively learns the correlation between channels, highlights important channels and suppresses secondary channels. The spatial attention module (SAM) learns key areas in space and improves the perception of areas of interest. CBAM can not only improve the feature representation ability of the model, but also help to capture the details and context information in the image.

2.5. The proposed SE-CBAM-ResNet101 model

Figure 9 presents the architecture of the proposed SE-CBAM-ResNet101 model. The proposed SE-CBAM-ResNet101 model adds a CBAM attention mechanism module and an SE module. The CBAM attention mechanism module is added after the last average pooling layer of ResNet101 and the SE module is added after the CBAM module. The advantage of adding these two attention

mechanism modules to ResNet101 is that both the SE module and the CBAM module can perform multi-layer feature fusion. The SE module can perform attention adjustment on the output feature map while the CBAM module performs feature fusion on different modules or different layers of the convolutional neural network. This multi-layer feature fusion can capture different levels of detailed information and contextual relations and enhance the expressive ability of the model. The combination of the two may obtain a more comprehensive and refined feature representation.

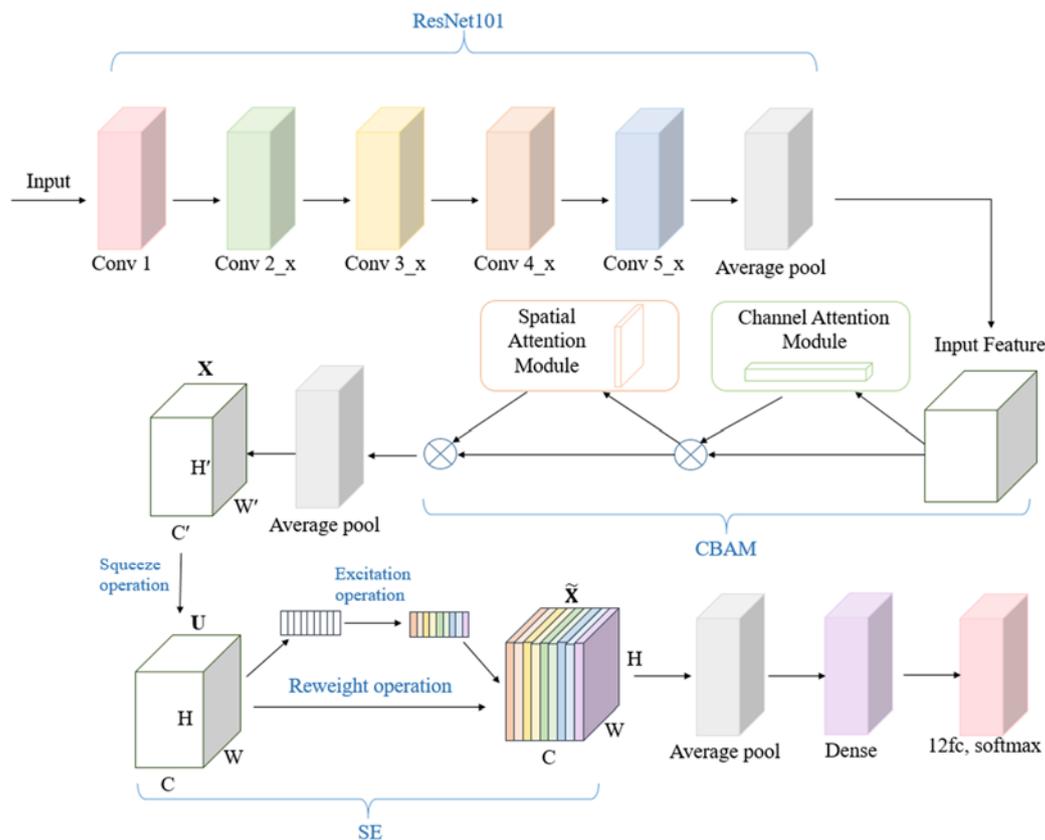


Figure 9. Architecture of the proposed SE-CBAM-ResNet101.

2.6. Evaluation of model performance

A confusion matrix reports the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) which is commonly used to evaluate the classification performance of models. TP represents the number of positive categories that are correctly classified as positive; FP represents the number of negative categories that are incorrectly classified as positive; TN represents the number of negative categories that are correctly classified as negative; and FN represents the number of positive categories that are incorrectly classified as negative. We selected the following performance indicators: accuracy, precision, recall and F1-score to evaluate the performance of each model. In addition, one more metric PR-AUC is calculated for the unbalanced dataset used in this study.

1) Accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

2) Precision

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3) Recall

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4) F1-score

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

3. Results

In the experiment, five-fold cross validation is used to evaluate the performance of each model. Table 2 shows the parameter setting for the experiments. The input size is 224×224 ; batch size at 8; epochs count of 50; optimizer of Adam, learning rate of 0.00001 with cross-entropy loss function. The equipment used in the experiment is not the latest, we have an Intel(R) Core(TM) i7-10700F 2.81 GHz CPU, NVIDIA GeForce RTX 2070 8 G GPU using Python 3.7 [Python Software Foundation, Fredericksburg, Virginia, USA] which contains keras 2.6 and tensorflow 2.6.

Table 2. Parameter setting for ResNet101.

Parameter	42.68M
Input size	224*224
Batch size	8
Epoch	50
Learning rate	0.00001
Classes	12

As presented in Tables 3 and 4, the parameter, accuracy, specificity, sensitivity and F1-score of AlexNet were 62M, 71.28%, 73.65%, 71.28% and 71.35%, respectively. The parameter, accuracy, specificity, sensitivity and F1-score of InceptionV3 were 79.53%, 74.84%, 79.65% and 76.30%, respectively. The parameter, accuracy, specificity, sensitivity and F1-score of ResNet50 were 75.20%, 76.25%, 75.27% and 74.79%, respectively. The parameter, accuracy, specificity, sensitivity and F1-score of ResNet101 were 75.26%, 77.10%, 75.28% and 75.33%, respectively. Figure 10 displays the accuracy and model size for each model. Figure 11 shows the boxplots for all models. From the boxplot, obviously, the accuracy from AlexNet is the lowest while InceptionV3 achieves the highest accuracy of 79.53%. Although, the accuracy comes from InceptionV3 is the highest, the precision from ResNet101 is the highest. ResNet has a simpler, single scale processing unit. Inception focuses on computational cost while ResNet emphasizes on computational accuracy. Therefore, we select ResNet for the further experiments in the following sections.

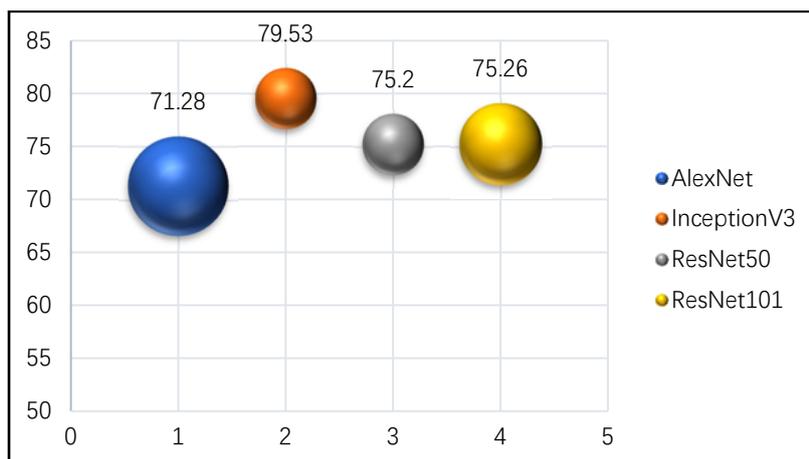


Figure 10. The accuracy and model sizes.

Table 3. Model size.

	AlexNet	InceptionV3	ResNet50	ResNet101
Parameter	62 M	23 M	23 M	42 M

Table 4. Model performance.

Model	Index (%)	Fold					Average \pm SD (%)
		1	2	3	4	5	
AlexNet	Accuracy	81.40	65.11	73.26	71.51	65.12	71.28 \pm 6.75
	Precision	82.55	67.18	78.33	73.62	66.57	73.65 \pm 6.95
	Recall	81.40	65.12	73.26	71.51	65.12	71.28 \pm 6.75
	F1-score	80.93	65.00	73.86	72.03	64.91	71.35 \pm 6.71
InceptionV3	Accuracy	90.70	81.40	76.74	70.30	78.49	79.53 \pm 7.46
	Precision	91.29	76.08	72.23	62.17	72.42	74.84 \pm 10.55
	Recall	90.70	81.40	76.74	70.93	78.49	79.65 \pm 7.26
	F1-score	90.72	77.85	73.36	65.28	74.27	76.30 \pm 9.28
ResNet50	Accuracy	79.65	68.03	75.40	77.33	75.58	75.20 \pm 4.36
	Precision	80.44	68.93	75.08	78.83	77.97	76.25 \pm 4.53
	Recall	79.65	68.02	75.78	77.33	75.58	75.27 \pm 4.37
	F1-score	79.25	67.27	74.90	76.93	75.59	74.79 \pm 4.52
ResNet101	Accuracy	80.81	65.12	72.35	78.49	79.51	75.26 \pm 6.53
	Precision	83.56	67.18	70.54	81.90	82.31	77.10 \pm 7.64
	Recall	80.81	65.12	72.35	78.49	79.65	75.28 \pm 6.55
	F1-score	80.67	65.00	72.42	78.75	79.83	75.33 \pm 6.63

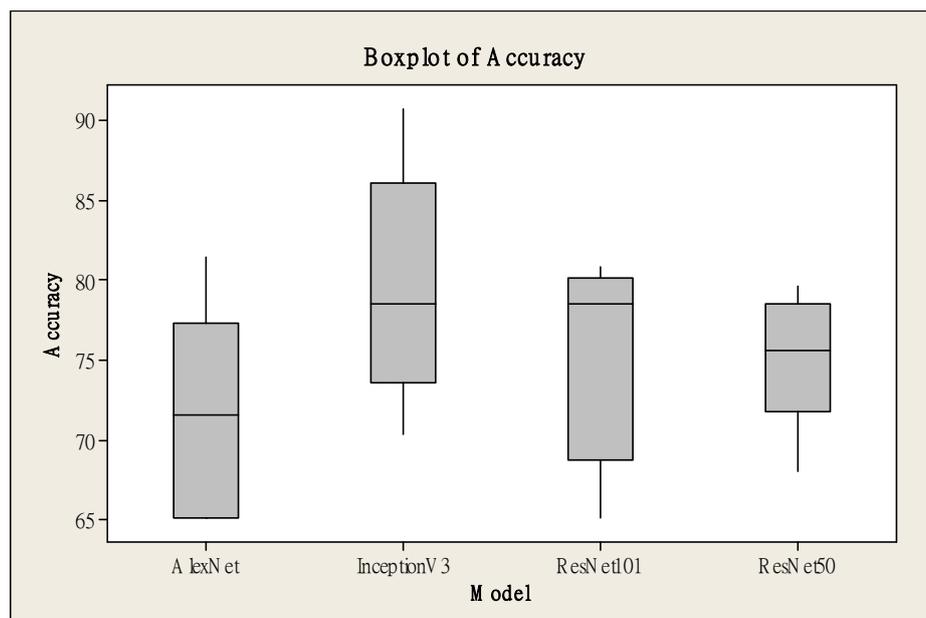


Figure 11. Accuracy boxplot for models.

4. Ablation experiments

To demonstrate the effectiveness of the proposed model, we have competed several experiments to observe if there is synergy between the different changes by adding squeeze-and-excitation (SE) module, adding convolutional block attention module (CBAM) or using Bayesian optimization (BO). Table 5 presents the result of ablation experiments.

Table 5. Comparisons on ablation experiment (%).

	ResNet101	SE	CBAM	BO	Accuracy	Precision	Recall	F1-score	PR-AUC	Size (M)	Ratio*
1	✓	✓			74.70	70.11	70.53	66.39	79.59	43.10	1.733
2	✓		✓		91.17	90.75	90.59	90.46	94.78	51.07	1.785
3	✓			✓	97.64	97.99	97.86	97.82	99.72	42.68	2.288
4	✓	✓	✓		83.52	61.79	72.56	66.03	83.13	51.60	1.619
5	✓		✓	✓	95.88	95.30	94.89	95.00	98.76	51.07	1.877
6	✓	✓		✓	77.64	84.09	73.08	74.17	84.57	43.10	1.801
7	✓	✓	✓	✓	83.52	74.57	78.38	75.57	82.58	51.60	1.619

*Note: Ratio = Accuracy divided by parameter quantity (M).

The best performance for accuracy, precision, recall, F1-score and PR-AUC are coming from ResNet101 model with Bayesian optimization (Experiment #3 in Table 5) which are 97.64%, 97.99%, 97.86%, 97.82%, 99.72%, respectively. The parameter setting for the proposed SE-ResNet101 (Experiment #1), CBAM-ResNet101 (Experiment #2), SE-CBAM-ResNet101 (Experiment #4) models are the same. The parameter setting for the experiment #3, #5, #6, and #7 Bayesian optimization is batch size from 8 to 32, epoch 30 to 70 and learning rate ranges 1e-5 to 1e-2. The best performance occurs at batch size 18; epoch count of 56; learning rate of 1.114e-05. From the result in Table 5, compared with adding convolutional block attention module (CBAM), adding squeeze-and-

excitation (SE) module in ResNet101 model does not have a positive effect on the performances. In addition, ratio of accuracy to parameter quantity is calculated. Obviously, with the highest accuracy 97.64% and the smallest model size 42.68 M, the highest ratio 2.288 occurs in experiment #3.

In addition to ResNet101, we have conducted more experiments on VGG and DenseNet. As shown in Table 6, VGG16 and VGG19 with and without SE or CBAM are compared. The best accuracy is from VGG16 with SE. The accuracy achieves 97.06% and the PR-AUC is as high as 98.64% (Experiment #1 in Table 6). While the accuracy of Experiment #4 VGG19 with CBAM is 97.05% which is close to the highest accuracy in VGG16-SE model, the PR-ACU is lower than 98%. The accuracy from DenseNet121-SE is 97.06% while the PR-AUC from DenseNet121 with channel attention module achieves 99.62% (Table 7). Comparing the performance in Tables 5–7, the overall highest accuracy 97.64% and the PR-AUC 99.72% come from ResNet101 model with Bayesian optimization (Experiment #3 in Table 5).

Table 6. VGG experiments (%).

	VGG16	VGG19	SE	CBAM	Accuracy	Precision	Recall	F1-score	PR-AUC
1	✓		✓		97.06	96.29	96.64	97.27	98.64
2	✓			✓	94.11	94.46	94.11	93.95	98.29
3		✓	✓		91.18	88.45	88.68	90.36	95.71
4		✓		✓	97.05	97.28	97.05	96.86	97.35

Table 7. DenseNet121 experiments (%).

	DenseNet121	SE	CAM*	Accuracy	Precision	Recall	F1-score	PR-AUC
1	✓	✓		97.06	97.17	97.06	97.02	99.57
2	✓		✓	95.88	96.08	95.88	95.88	99.62

*Note: Channel Attention Module.

5. Conclusions

Chinese herbal medicine flowers have considerable economic value. Flowers are easily confused due to their similar appearance. The SE-ResNet101 model proposed in this study is used for image recognition and classification on traditional Chinese medicine flowers by using ImageNet for pre-training and add SENet to the trained ResNet101. The research results show that the best performance for accuracy, precision, recall, F1-score and PR-AUC are coming from ResNet101 model with Bayesian optimization which are 97.64%, 97.99%, 97.86%, 97.82%, 99.72%, respectively. Compared with four models including AlexNet, InceptionV3, ResNet50 and ResNet101, our proposed ResNet101 model has achieved the best performance while maintaining the total number of parameters favorable. Since the dataset is limited to twelve common Chinese medicinal flowers, the comprehensive review of Chinese medicinal flowers is suggested in the future research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors gratefully acknowledge the financial support of the Ministry of Science and Technology of Taiwan through the grant MOST 111-2221-E-167-007-MY3.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. H. Yuan, S. Jiang, Y. Liu, M. Daniyal, Y. Jian, C. Peng, et al., The flower head of *Chrysanthemum morifolium* Ramat.(Juhua): A paradigm of flowers serving as Chinese dietary herbal medicine, *J. Ethnopharmacol.*, **261** (2020), 113043. <https://doi.org/10.1016/j.jep.2020.113043>
2. Y. Xu, G. Wen, Y. Hu, M. Luo, D. Dai, Y. Zhuang, et al., Multiple attentional pyramid networks for Chinese herbal recognition, *Pattern Recognit.*, **110** (2021), 107558. <https://doi.org/10.1016/j.patcog.2020.107558>
3. F. Jiang, Y. Lu, Y. Chen, D. Cai, G. Li, Image recognition of four rice leaf diseases based on deep learning and support vector machine, *Comput. Electron. Agric.*, **179** (2020), 105824. <https://doi.org/10.1016/j.compag.2020.105824>
4. P. Kumari, B. Bhargava, Phytochemicals from edible flowers: Opening a new arena for healthy lifestyle, *J. Funct. Foods*, **78** (2021), 104375. <https://doi.org/10.1016/j.jff.2021.104375>
5. T. Lv, R. Teng, Q. Shao, H. Wang, W. Zhang, M. Li, et al., DNA barcodes for the identification of *Anoectochilus roxburghii* and its adulterants, *Planta*, **242** (2015), 1167–1174. <https://doi.org/10.1007/s00425-015-2353-x>
6. Y. Chen, J. Huang, Z. Q. Yeap, X. Zhang, S. Wu, C. H. Ng, et al., Rapid authentication and identification of different types of *A. roxburghii* by Tri-step FT-IR spectroscopy, *Spectrochim. Acta, Part A*, **199** (2018), 271–282. <https://doi.org/10.1016/j.saa.2018.03.061>
7. A. Jahanbakhshi, Y. Abbaspour-Gilandeh, K. Heidarbeigi, M. Momeny, Detection of fraud in ginger powder using an automatic sorting system based on image processing technique and deep learning, *Comput. Biol. Med.*, **136** (2021), 104764. <https://doi.org/10.1016/j.compbiomed.2021.104764>
8. K. IlBae, J. Park, J. Lee, Y. Lee, C. Lim, Flower classification with modified multimodal convolutional neural networks, *Expert Syst. Appl.*, **159** (2020), 113455. <https://doi.org/10.1016/j.eswa.2020.113455>
9. Q. Chai, J. Zeng, D. Lin, X. Li, J. Huang, W. Wang, Improved 1D convolutional neural network adapted to near-infrared spectroscopy for rapid discrimination of *Anoectochilus roxburghii* and its counterfeits, *J. Pharm. Biomed. Anal.*, **199** (2021), 114035. <https://doi.org/10.1016/j.jpba.2021.114035>
10. Y. Xu, G. Wen, Y. Hu, M. Luo, D. Dai, Y. Zhuang, et al., Multiple attentional pyramid networks for Chinese herbal recognition, *Pattern Recognit.*, **110** (2021), 107558. <https://doi.org/10.1016/j.patcog.2020.107558>
11. M. L. Huang, Y. X. Xu, Chinese medicinal blossom-dataset, Mendeley Data, V1, 2021. <https://doi.org/10.17632/r3z6vp396m.1>

12. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **2016** (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
13. K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
14. W. Zhou, H. Wang, Z. Wan, Ore image classification based on improved CNN, *Comput. Electr. Eng.*, **99** (2022), 107819. <https://doi.org/10.1016/j.compeleceng.2022.107819>
15. X. Zhao, K. Li, Y. Li, J. Ma, L. Zhang, Identification method of vegetable diseases based on transfer learning and attention mechanism, *Comput. Electron. Agric.*, **193** (2022), 106703. <https://doi.org/10.1016/j.compag.2022.106703>
16. A. Ma, Y. Wan, Y. Zhong, J. Wang, L. Zhang, SceneNet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search, *ISPRS J. Photogramm. Remote Sens.*, **172** (2021), 171–188. <https://doi.org/10.1016/j.isprs.2020.11.025>
17. Y. Wan, Y. Zhong, A. Ma, J. Wang, L. Zhang, E2SCNet: Efficient multiobjective evolutionary automatic search for remote sensing image scene classification network architecture, *IEEE Trans. Neural Networks Learn. Syst.*, **2022** (2022). <https://doi.org/10.1109/TNNLS.2022.3220699>
18. D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, D. Li, An efficient and lightweight convolutional neural network for remote sensing image scene classification, *Sensors*, **20** (2020), 1999. <https://doi.org/10.3390/s20071999>
19. Z. Wu, F. Jiang, R. Cao, Research on recognition method of leaf diseases of woody fruit plants based on transfer learning, *Sci. Rep.*, **12** (2022), 1538.
20. Y. Liu, Y. Sun, B. Xue, M. Zhang, G. G. Yen, K. C. Tan, A survey on evolutionary neural architecture search, *IEEE Trans. Neural Networks Learn. Syst.*, **34** (2023), 550–570. <https://doi.org/10.1109/TNNLS.2021.3100554>
21. H. Li, Skin burns degree determined by computer image processing method, *Phys. Procedia*, **33** (2012), 758–764. <https://doi.org/10.1016/j.phpro.2012.05.132>
22. H. He, X. Huang, Y. Song, Z. Zhang, M. Wang, B. Chen, et al., An insulator self-blast detection method based on YOLOv4 with aerial images, *Energy Rep.*, **8** (2022), 448–454. <https://doi.org/10.1016/j.egy.2021.11.115>
23. L. Yang, H. Yu, Y. Cheng, S. Mei, Y. Duan, D. Li, et al., A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture, *Comput. Electron. Agric.*, **187** (2021), 106316. <https://doi.org/10.1016/j.compag.2021.106316>
24. R. Zhang, J. Zhao, H. Xie, T. Wang, G. Chen, G. Zhang, et al., Automatic diagnosis for aggressive posterior retinopathy of prematurity via deep attentive convolutional neural network, *Expert Syst. Appl.*, **187** (2021), 115843. <https://doi.org/10.1016/j.eswa.2021.115843>
25. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-Excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 2011–2023. <https://doi.org/10.1109/TPAMI.2019.291337>
26. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

